DOCUMENT RESUME

ED 473 939 TM 034 788

AUTHOR	Claycomb, Carla; Kysilko, David
TITLE	The Purposes & Elements of Effective Assessment Systems. The Policy Framework.
INSTITUTION	National Association of State Boards of Education, Alexandria, VA.
PUB DATE	2000-03-00
NOTE	7p.
PUB TYPE	Information Analyses (070) -- Journal Articles (080)
JOURNAL CIT	State Education Standard; v1 n2 p6-11 Spr 2000
EDRS PRICE	EDRS Price MF01 Plus Postage. PC Not Available from EDRS
DESCRIPTORS	*Educational Change; Elementary Secondary Education; Large Scale Assessment; State Programs; *State Standards; Student Evaluation; *Testing Programs

ABSTRACT

Nearly every state is engaged in standards-based reform, as seen by the development of statewide academic goals and guidelines in core subjects and the subsequent new state tests that will measure students' progress toward achieving these goals. A Study Group on Statewide Assessment Systems formed by the National Association of State Boards of Education (NASBE) explored issues related to standards-based reform and the assessment that supports it. This article reviews the Study Group's conception of the purposes of state assessments and uses these to build a framework for the essential characteristics of effective assessment systems. The ultimate goal of state standards and assessments is not to sort students into "haves" and "have nots" of knowledge, but to enable all students to meet the standards. An effective assessment system: (1) is aligned with rigorous state standards; (2) addresses specific goals and purposes; (3) balances validity, reliability, and efficiency; (4) inform instruction and providing for consequences; (5) has mechanisms to encourage schools and districts to align their instruction and evaluation with the state system; and (6) has a clearly articulated relationship with national measures of student performance. The NASBE Study Group was emphatic that state boards need to monitor the implementation of standards and assessment policies to ensure that their policy vision is realized. (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY

B. L. Welburn

2A TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

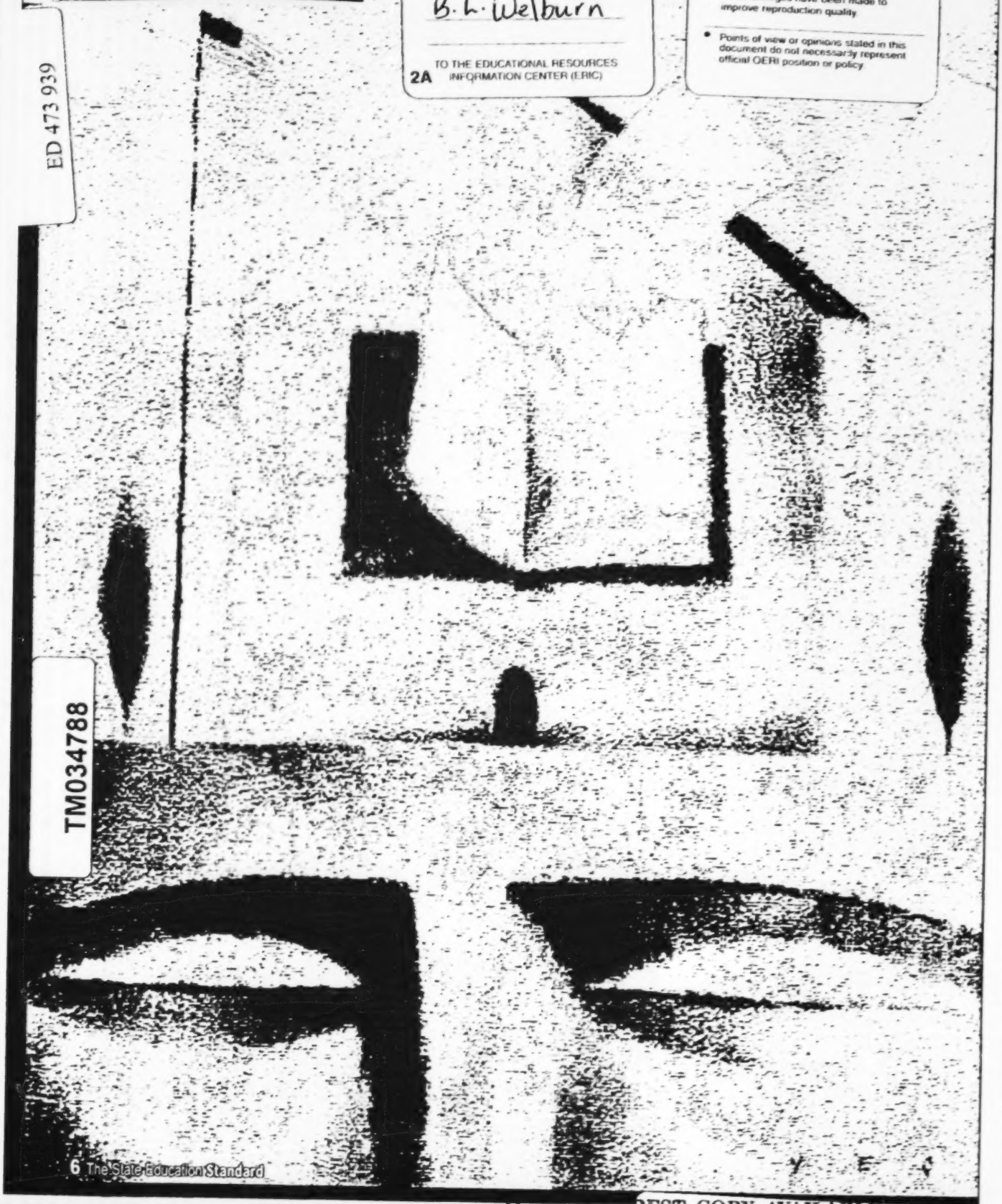
☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

ED 473 939

TM034788



BEST COPY AVAILABLE

The Purposes & Elements of Effective Assessment Systems

by Carla Claycomb and David Kysilko

For many years, governors, legislatures, and state boards of education have struggled to find the most effective role for the state in improving student achievement and providing oversight for the education system. Today, nearly every state is engaged in standards-based reform, as seen by the development of statewide academic goals and guidelines in core subjects and the subsequent new state tests that will measure students' progress toward achieving these goals. This has proved to be a complex, lengthy, and at times frustrating and controversial undertaking. In focusing on the testing aspects of this effort, NASBE's 1997 Study Group on Statewide Assessment Systems began with the most fundamental questions: Does standards-based reform make sense? Is this the best strategy states can use to ensure that all students are well-prepared for the future?[1]

The Study Group clearly affirmed the importance of state standards and noted that a host of issues had combined to make standards, in the words of a previous NASBE report, "one of the most powerful options for school reform." Moreover, despite some recent rough patches in their implementation, including various grassroots backlashes and Secretary Riley's call for

BEST COPY AVAILABLE

a "midcourse correction," the vast majority of state board members remain committed to standards. The Study Group laid out its reasons for supporting this reform strategy:

- Standards focus on outcomes, which in turn provide flexibility to local schools and districts in how to achieve those outcomes;
- Standards provide for uniform goals and a uniform measurement tool in an age of increased mobility and wide variations in local grading practices;
- Standards are developed through an open process and are publicly available for any teacher, parent, or student to see;
- Standards are set at high levels so that students are expected to learn challenging material, analyze information, and communicate results; and
- Standards apply to all students, not just the college bound.

This article, written by NASBE staff members who facilitated the work of the committee, is based on the Study Group's recommendations. It reviews the Study Group's conception of the purposes of state assessments and uses these to build a framework for the essential characteristics of effective assessment systems. It also includes updated information to reflect recent events in the world of assessment policy.

The ultimate goal of state standards and assessments is not to sort students into "haves" and "have nots" of knowledge. The goal is to enable all students to meet the standards.

The Purposes of Assessments in Relation to Standards

If standards provide the skeleton or framework on which states can build their academic system, assessments are the muscles that bring the system to life. By now all the states and other jurisdictions that are developing or have developed standards are also aligning their state assessments with the new standards. This is critical, for the nature of these new assessment systems may well determine whether the standards truly drive teaching, learning, equity, and accountability in the state; whether they are merely vaguely remembered pronouncements that gather dust in districts' central offices; or whether they are denounced by parents, teachers, and students alike as central cogs in a factory-model education system the narrows the curriculum, inhibits creativity, and drives low-achieving students out of schools. Whatever the final assessment system in a state looks like, policymakers initially need to step back and understand what their purposes for the assessments are and then regularly revisit this discussion over the years to ensure that the assessments as implemented are indeed fulfilling those purposes. Some of the underlying purposes are as follows:

- **Assessments help ensure that standards are taken seriously.** When the content of state assessments is closely aligned with that of the standards, this is a clear signal to teachers, administrators, students, and the public that the state is serious about its academic standards. Basically, as all teachers and students know, if "it's on the test," it must be important. Aligning assessments with standards is critical for another reason, as well: if material is covered on the assessment but is not in the standards, it is inherently unfair to students and teachers alike. In such cases, the material may not have been covered in class because, again, no one knew it was important.

  Not only must content be both in the standards and on the assessment, but the assessment must mean something: there must be "stakes" attached to the assessments at some point. For students that point is most likely to be in terms of a diploma; that is, a student's level of achievement on state assessments could be indicated on the diploma, or meeting the state's academic standards could be one requirement for graduation. For districts and schools, it means student scores on state assessments should be one component of the state's accountability plan.
- **Standards and assessments guide teaching and learning.** Clearly articulated content standards and the examples used in performance standards or curriculum frameworks to illustrate the knowledge and skills embodied in the standards are intended to guide teachers toward the state's vision of essential teaching and learning. Of course, teachers will vary in how closely they adhere to the examples in creating their lessons. The point is, when standards are taken seriously, all teachers will share a vision—shaped by a consensus of the best thinking in the state—of the knowledge and skills students should acquire.
- **Assessments help individual students meet the standards.** The ultimate goal of state standards and assessments is not to sort students into the "haves" and "have nots" of knowledge. The goal is to enable all students to meet the standards. In this sense, results on state assessments should be part of a support system that can signal when a student is in trouble

BEST COPY AVAILABLE

academically and needs extra help or special instruction. At the same time, teachers should be able to use the standards in developing their own classroom assessments to help identify students who may need more instruction or to inform a teacher when it is time to adjust his or her curriculum or instructional strategies. If states are serious about having all students meet the standards, they will provide assistance to districts and schools for this remediation.

- **Assessments help policymakers ensure that all students have access to a sound education.** Policymakers want to assure the public that all students, no matter where they live in the state, have a reasonable opportunity to get a sound education. As indicated above, having assessments as one part of the state's accountability system helps ensure that the standards are taken seriously. Just as important is using the accountability system to verify that all schools and districts are providing a "reasonable opportunity." This is a basic state responsibility—indeed, it is often specifically addressed in state constitutions. Some have expressed concern that standards and high-stakes assessments will only mean that more of those students who were having trouble in the past will not graduate in the future. Instead, the Study Group believes that standards and assessments, when part of an overall evaluation and accountability system, can help provide the state with the information and the mechanisms it needs to bring a solid education within the reach of every student.

Elements of Effective Assessment Systems

Over the years states have developed a wide range of assessment systems with different purposes and consequences, reflecting the unique content and character of each state's education system and traditions. Such diversity is a unique and vital aspect of the nation's heritage of state and local control of education. However, the Study Group concluded that there are general characteristics that highly effective state assessment systems share. A brief look at these essential elements follows.

(1) An Effective State Assessment System is Aligned with Rigorous State Standards.

Standards need at least two characteristics to be effective. First, they must be rigorous and detailed enough to guide curriculum and assessment development. Second, standards must be put in force via a closely aligned assessment system. Many states have developed standards that are not rigorous or specific enough to meaningfully direct assessment development. In these states, it is likely to be the content and structure of their assessments, rather than the content of their standards, that drive school reform.

(2) An Effective State Assessment System Addresses Specific Goals and Purposes.

Whereas standards define what to test, assessment goals define who to test, when to test, and what type of test to use. As indicated above, the Study Group felt strongly that a primary purpose for any state assessment system should be to improve teaching and learning. However, tests to improve teaching and learning can take many different forms and address a variety of complementary goals and purposes. For example, tests to inform student remediation need to be given to every student in the testing grade, constructed to give rich information to teachers, given sufficiently early enough in the year to inform instruction, and designed for efficient scoring so that results can be returned to the classroom as soon as possible. Tests to evaluate schools or districts can be given to a sample of students and need not be given every year, which saves money and instructional time.

Aligning state assessments with state goals and purposes makes good financial sense. Assessments that are implemented without clear goals may measure variables that are unnecessary or may measure necessary variables in inefficient ways. For example, measuring every student's writing ability to inform district curricular improvement in writing is "overkill"; only a sample of students need to be tested for that purpose. But if a state goal is to ensure that all graduates can write a well-formed paragraph, every student must be tested.

(3) An Effective State Assessment System Balances Validity, Reliability, and Efficiency.

For assessment results to mean anything, the test must be both valid (meaning that it measures what it purports to measure) and reliable (meaning, among other things, that students would get similar results if tested again). Statistically, it is impossible for a test to be completely valid and reliable, but even making them "sufficiently" valid and reliable to satisfy stakeholders while at the same time being fair and equitable is a challenge.

Fundamentally, those who write assessments have two types of questions to choose from: multiple- choice and performance. In general, multiple-choice tests are very reliable provided they include a sufficiently large number of items. Multiple-choice tests are also efficient in terms of testing time and cost. But they do not resemble good instruction. Furthermore, multiple-choice tests essentially measure the ability to choose the correct answer, which is not a skill often required by adults in the "real world." For these reasons, multiple-choice tests have been criticized for having low validity.

Performance assessments are more likely to measure skills that policymakers want students to develop, such as the ability to write a persuasive paragraph or to draw inferences from a science experiment. Constructed response questions, which require students to write short-essay responses or solve problems and show their work, are currently the most popular performance items used on state tests, though some states have also worked with more extended performance tasks and portfolios. However, experience in Kentucky, Maryland, and Vermont, among other states, indicates that the various types of performance assessment differ in their validity and reliability. In general, constructed response items administered in a traditional testing situation are more likely than other types of performance assessment to be scored accurately enough to produce reliable test results. Of course, they are less efficient than multiple-choice questions because of the greater testing time per item, and because more time and money are required to score responses accurately and consistently.

Because no one test can "do it all," the Study Group recommended that states consider ensuring fairness and curricular coverage and depth by implementing an assessment *system* with more than one test that balances performance assessment items with multiple-choice. The National Assessment of Educational Progress (NAEP), the nation's primary measure of what students

The Study Group recommended that states consider...implementing an assessment system with more than one test that balances performance assessment items with multiple-choice.

know and can do on a state and national level, is a good example of a test that combines multiple-choice and performance items and has proven both valid and reliable for use at the state and national, rather than individual, level.

(4) An Effective State Assessment System Informs Instruction and Has Consequences.

Two primary reasons for developing a state assessment system are (1) to inform student instruction by pointing out student learning gaps and students needing acceleration, and (2) to assign rewards and sanctions to students, teachers, schools, or districts based upon assessment outcomes.

Informing Instruction. The Study Group asserts that once states establish standards, they have both a moral and legal obligation to provide all students with the instruction they need to attain them. Assessments are instrumental to diagnose deficiencies and to help teachers target instruction for individual students. Especially in elementary grades, a state test can be a powerful tool to focus attention and resources on helping students, schools, and districts meet standards. Unfortunately, not all states are including remediation in their assessment agendas: while 40 states require or have plans to require remediation for students not meeting standards, only 29 back the requirement with funds, and only 13 require and fund remediation in the four core subject areas.[3]

Helping all students reach their capacity also requires that teachers receive high-quality information about students who score well on state assessments. Once states establish a system to encourage all students to achieve high standards, they have an obligation to provide high-achieving students with the instruction they need to develop to their full capacity. If a student's performance on a state assessment points out subject areas in which performance exceeded state expectations, there should be mechanisms in place, based upon assessment results and state standards, to encourage that student to continue achieving.

Having consequences. Although providing remediation and acceleration to particular students is necessary for an effective accountability system, it is not sufficient. To have the greatest impact on teaching and learning, an assessment system also needs to have consequences. Making assessments count as one component of an accountability plan is important because studies show that high school students perform better on assessments that have consequences for them. Teachers are also more likely to take test results seriously and more inclined to change their instructional practices to reflect the content and format of a test if the test has consequences.[4]

Currently, 28 states tie or plan to tie the award of high school diplomas to achievement of their standards.[5] In these, assessments are commonly used alone, rather than as one component of a multifaceted evaluation system. Denying diplomas to students based on test scores when students are otherwise qualified to graduate means that students who do well in school but perform poorly on the state assessment may be unfairly penalized by a one-shot evaluation of their accumulated school work. Consequently, the Study Group asserts that either: (1) high-stakes tests should be only one of several components that are considered when making decisions about student promotions and graduation; or (2) state assessments should be used as high-stakes measures in ways other than graduation or promotion, such as for conferring diplomas with endorsements.

More common than using state assessments to assign consequences to individual students is using assessment results for school or district accountability. However, the same caution applies to schools and to districts as to individual students: no school should be penalized on the basis of test results alone. Rather, test results should be one component of an evaluation system that may also include variables such as dropout rates, teacher attrition rates, expenditure information, and attendance rates.

The Study Group strongly recommended that states guarantee that students have the remediation or acceleration necessary to learn to the highest standards. One state board that is currently taking this issue seriously is Maryland, where the board continues to fight for its "K-12 Intervention" plan, a proposal to create a $49 million remedial education program to help students keep up with the learning expectations set out in the state standards and assessments. The board has signaled to the governor and legislature that if there is not sufficient funding for this program, they will have to consider alternatives that basically water down the standards by lowering the passing grade or eliminating passing state tests as a requirement for graduation.[6]

(5) An Effective State Assessment System Has Mechanisms to Encourage Schools and Districts to Align Their Instruction and Evaluation with the State System.

Standards and assessments should serve as a catalyst to direct local reforms in improving teaching and learning. Most

BEST COPY AVAILABLE

states seem to assume that instituting standards and assessments will prompt school districts to align their practices with them. Research suggests that such an assumption may be warranted for forward-thinking districts that have the capacity to tackle change head-on. Schools that lack the will or capacity to change, however, are unlikely to benefit from state mandates alone.

Explicit programs linking state standards and assessments with district and classroom teaching and assessment practices are critical to promote student achievement in all schools and districts. One effective way to impact classroom teaching and assessment practices is to include performance items on the state assessment. In general, teachers are more likely to incorporate the goals and skills of a performance assessment in their instruction, but they need help to do so effectively.

(6) An Effective State Assessment System Has a Clearly Articulated Relationship with National Measures of Student Performance.

As states invest time and resources in developing a comprehensive assessment system, they must ensure they can calibrate assessment results with the performance of students across the country. Currently, most states that have articulated standards have designed assessment programs unique to their state. Because of this, states cannot compare their performance or that of individual students, schools, or districts with those of other states and have no quantifiable way of demonstrating that their students perform to national standards. The Study Group contended that being serious about standards and assessments means ensuring that standards define, and assessments measure, world class performance.

The best national benchmarks available for standards and assessments are the NAEP tests and NAEP subject frameworks. These frameworks and assessments reflect a general consensus about important competencies and expected levels of performance at key ages and in core subjects. Although North Carolina is a notable exception, most state standards and assessments are not aligned with the NAEP subject frameworks, tests, or standards. As a result, states that participate in NAEP may have widely differing results on national and state indicators. For example, whereas 35 percent of Wisconsin fourth- grade students were "proficient" on the 1996 NAEP math assessment, 88 percent of fourth graders were proficient on Wisconsin's own state exam. In Louisiana, seven percent were "proficient" on NAEP, whereas 80 percent of fourth graders were "proficient" according to the state assessment. Many states reporting results from different programs run into the problem of seemingly inconsistent results because the two programs have different frameworks and performance standards. Reporting such widely divergent results can be confusing to parents and the general public and send mixed messages to teachers and administrators concerning what content and skills to teach.

Conclusion

When state policymakers begin building an effective state assessment system, they are embarking on a complex and time-consuming process. This is because, perhaps even more than most other policy issues, the process of building a standards-led education system does not end for policymakers when the ink on the policy dries. NASBE's Study Group was emphatic that state boards need to monitor the implementation of standards and assessment policies to ensure that their policy vision is realized. This means, minimally, that substantial conversations about standards and assessments should appear with regular frequency on the board agenda. Preferably, these conversations should compare progress in assessment development and implementation with the criteria listed above to ensure that the assessment is as effective as possible. Best of all, policymakers should be clear in defining the type of assessments they envision. State policymakers that proactively and specifically delineate key characteristics of their state assessments are in the best position to ensure that they drive a system that fosters high achievement for all students entrusted to their care.

Carla Claycomb is an education researcher and writer. David Kysilko is editor of the State Education Standard. *In 1997 they staffed NASBE's Study Group on Statewide Assessment Systems.*

---

1. Study Group on Statewide Assessment Systems, *The Full Measure* (Alexandria, VA: The National Association of State Boards of Education, 1997).
2. J. Kendall and R. Marzano, *The Fall and Rise of Standards-Based Education* (Alexandria, VA: The National Association of State Boards of Education, 1996).
3. H. Glidden, *Making Standards Matter 1999* (American Federation of Teachers, 1999).
4. R. Truby, unpublished presentation to the NASBE Study Group on Statewide Assessment Systems, 15 March 1997.
5. H. Glidden, *Making Standards Matter 1999.*
6. "Fund the Standards," *Washington Post*, 8 March 2000.

Even more than most other policy issues, the process of building a standards-led education system does not end for policymakers when the ink on the policy dries.

BEST COPY AVAILABLE

TM034788

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

ERIC®

NOTICE

Reproduction Basis

☒ This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

☐ This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").